Ole Husby,
Linking in Union Catalogs

from / *aus*:

Union Catalogs at the Crossroad

Edited by

Andrew Lass and Richard E. Quandt

pp. / *S*. 89-100

Impressum

Bibliographic information published by Die Deutsche Bibliothek

Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.ddb.de.

Bibliografische Information Der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.ddb.de abrufbar.

This publication is also openly accessible at the publisher's website. Die Deutsche Bibliothek has archived the electronic publication, which is now permanently available on the archive server of Die Deutsche Bibliothek.

Diese Publikation ist auf der Verlagswebsite ebenfalls open access *verfügbar. Die Deutsche Bibliothek hat die Netzpublikation archiviert. Diese ist dauerhaft auf dem Archivserver Der Deutschen Bibliothek verfügbar.*

Available open access / *open acess* verfügbar:

Hamburg University Press / Hamburg University Press
http://hup.rrz.uni-hamburg.de
Die Deutsche Bibliothek archive server / Archivserver Der Deutschen Bibliothek
http://deposit.ddb.de/

ISBN 3-937816-08-9 (print)

Table of Contents

Part 6 South African Union Catalogs

Chapter 4
Linking in Union Catalogs

Ole Husby

> Identifying and categorizing relations is a necessary requirement for the formal description that makes navigation possible in the bibliographic universe.
> – Knut Hegna

1 Introduction

Lately, we have seen a number of new developments of union catalogs, regarding both their content and technical implementations. New material types are making their way into these catalogs/databases, the most notable perhaps being the network documents, residing somewhere on the Internet instead of on a library shelf. It naturally follows that new 'document delivery' mechanisms are in demand, and that the notion of holdings needs to be redefined.

In addition, so-called digital libraries are emerging, more or less to complement or even to include the services offered by the library catalogs. A core element of these digital libraries is a technology for managing, expressing and navigating relationships. However, the same importance should be attributed to relationships in library catalogs and union catalogs, especially when available on the Web. This paper will discuss a number of aspects of bibliographic relations and their expression, here called linking. For the purpose of this paper, I will take the liberty of using the term 'union catalog' in the broadest possible meaning, not even trying to offer a definition.

Without jumping to conclusions, I will claim that there is a trend within many information services towards the use of more dynamic data models and technical solutions, allowing relationships to be investigated or synthesized as the information space emerges or is explored. A popular description of this is 'just-in-time' links, as opposed to 'just-in-case.'

A lack of linking facilities might lead to replication of data and cataloging effort, redundancies and inconsistencies in the data structures, and more cumbersome tasks for users to collect together items that belong together and to distinguish between items that do not. The basic need for this collecting and separating was especially needed in the card catalog, but is still of vital importance, at least in all the cases where we are burdened by the much-discussed information overload.

Linking allows iterative information seeking, where the selection of specific manifestations or the selection of the desired (appropriate) items should be separated from, and appear at a later stage than, the topical discovery processes.

2 Search Portals, Cross-Searching, Union Catalogs

Since numerous different search portals are being offered, we also have to consider whether the service expected from a union catalog is best obtained by the distributed search paradigm, as we are doing when using a standard network protocol like Z39.50 to synthezise virtual union catalogs. The other option is to stick to the 'real' union catalog in the physical sense, but now perhaps with the possibility of using other record collecting mechanisms, such as the harvesting protocol offered by the Open Archives Initiative (the OAI-PMH protocol).

While choosing models and technologies for cross-searching and unifying services is not the topic of this chapter, let me just point out that the needs for such efforts will not disappear, as users increasingly expect to find everything they need at the 'one-stop shop.' But there is certainly no single solution that should be recommended for all purposes.

Another apparent development is that the criteria for including diverse document categories in the union catalog need rethinking: "In which database does this record belong?" For electronic journals one might prefer

to have a separate e-journal database, for freely available network documents the subject portals could be suitable alternatives, and so on. This question will not be further treated here, but there is no single answer to this question either.

However, what in my opinion is important in this context is that the need for linking is apparently greater than ever before.

3 What is a Link?

Several definitions of links are available, among others:

- A link is an expression of a relation;
- A link is a connection from one page to another destination, such as another page, or a different location on the same page; and
- A link is underlined and blue.

As the latter two are, in my opinion, too strongly tied up with the World Wide Web way of thinking, I will choose the first one. With slightly different wording, we could say that a link is an instantiation of a relation—in the hypertext language, we could call it traversable or executable.

This definition next requires some discussion of relations. In general, a relation represents a meaningful connection between two or more entities. The concept of a relation can be rigorously defined in mathematical terms, which, however, I will not do, since in any case there does not seem to be too much variation in how we understand this term. Neither will I go into the typology of relations, but briefly mention that there are several ways of classifying them. One interesting aspect is to distinguish between

- relations that are a priori given by the nature of things;
- relations that are made up by us, and
- relations that are deduced from statistics.

4 A Link is an Expression of a Relation

As mentioned above, this is my preferred definition of a link. There are many ways of expressing it, and not all are hypertext links! Here are some quite different methods:

- Citing together;
- Explicitly stating in text ("See");
- Using controlled vocabularies;
- Data modeling (relational databases);
- Sharing metadata (identifiers, etc.);
- Linking in hypertext systems (blue, underlined ...).

In traditional thinking, linking has been seen as manifestations of relations between bibliographic records. This brings us next to the catalog.

5 Linking in Library Catalogs

In library catalogs, information about books, journals and other information entities are made available to the public. Us e of these systems, however, often demonstrates that it is not as easy and intuitive to locate the relevant information as we would like to think. A well-known problem is the failure of these databases to bring together objects that belong together, like translations of a given document into different languages, or the representation of the document in different media. There is a need for a conceptual model that captures the entities and relationships of primary concern to information retrieval.

Nearly all current catalogs describe (by the use of document surrogates) manifestations. This does not imply that the manifestations are the only entities present in the catalog, but rather that the descriptions of other more abstract entities are distributed in a different way. Multiple-item entities can be listed in one record, and information related to the expression and work entities will often be replicated in multiple records. One example is the MARC uniform title element, which acts as a sort of work title in the FRBR ("Functional Requirements for Bibliographic Records") model. However, this 'work' is not present as a distinct and identifiable entity in the catalog.

Maps for describing the entities and their interrelations may be constructed and integrated with the catalog in different ways, either as tightly integrated static parts of the records in the database, or else as a distinct and separate information service to be applied dynamically at runtime, as a separate link service. I will return to this topic later on.

6 The FRBR Model

The FRBR model has already been mentioned. And while most library catalogs are as yet quite unmarked by this important effort, it is evidently having a major impact on how system designers are modeling the next generation of bibliographic databases. An indication of the model's success is that it has been warmly welcomed far outside the community of theoretically inclined catalogers.

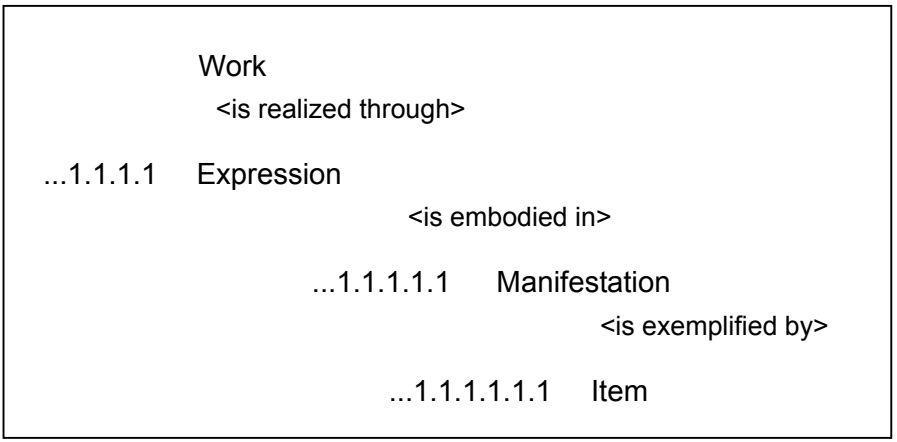

Figure 1. The Product Entities of the FRBR Model

"Functional Requirements for Bibliographic Records" (FRBR) is the title of a report from an IFLA study group, published in 1998. Briefly, FRBR presents a model for bibliographic data based on the entity-relationship data model. Three different groups of entities are defined:

- Group 1, the products of intellectual or artistic endeavor that are named or described in bibliographic records: work, expression, manifestation, and item;
- Group 2, entities responsible for the intellectual or artistic content, the physical production and dissemination, or the custodianship of such products: person and corporate body; and
- Group 3, entities that serve as the subjects of intellectual or artistic endeavour: concept, object, event, and place.

As an example, a work can be a novel, identified by a uniform title. Translations into different languages give the expressions, identified by title. A certain expression can be published on different media, giving the manifestations. Manifestations are often identified by ISBNs. The separate copies are then the items, identified (in the library space) by library codes and shelfmarks.

7 The FRBR Relations

The FRBR model further considers the following categories of relations, where some quite simple examples are given:

Between Work, Expression, Manifestation, and Item:

```
E2 <is translation of> E1
M1 <is manifestation of> E1
```

To Persons and Corporate Bodies:

```
P1 <is author of> W1
I1 <is owned by> C1
```

To Concept, Object, Event, Place:

```
W1 <is about> C1
```

Between Persons and Corporate Bodies:

```
P1 <often cites> P2
P3 <is often cited together with> P4
```

Between Concepts:

```
C1 <is subspecies of> C2
```

8 FRBRizing the Catalog

There is, in my opinion, potential for substantial improvement in library catalog linking, depending on the success of implementing the FRBR model in current and future library catalogs. So what can we do in practice? Here are some directions that should be investigated further:

- Rearrange bibliographic records (entities) according to FRBR, by changing the physical data model, or by extracting FRBR entities at runtime (just-in-time FRBR). This could mean having different types of records, corresponding to the product entity types in FRBR;
- Develop a sound framework for linking, if possible by using externally maintained link services or link databases;
- Choose record identifiers that support linking with as global a scope as possible. There are a number of standard identifier schemes for manifestation records, but hardly any for expressions or works. If we are concerned about interoperation and resource sharing, perhaps we should get together and invent new ones?
- Build maps, paths and navigational tools that guide the user in this new terrain. We have to take into account that user requirements and preferences differ strongly. Nowadays, many of us are used to having result sets sorted and arranged by search engines—often by non-intuitive and 'magic' clustering and ranking algorithms. We do not want magic systems (do we?), but comprehensible and accountable system behaviour.

It is not proven that the FRBR model is perfect, and the ideas above are far from easily realized. Experiments have shown that the automatic extraction of FRBR entities and relations may be a very tough task. But the time has not yet come to give in!

9 New Opportunities (and Needs)

Whichever definition of a link we might agree on, linking has become a new way of thinking that has emerged with the Web and hypertext. And whatever I might claim, links are perceived as “underlined and blue”... The new opportunities are welcomed and taken for granted by the users. It's now ‘up to the user to click.’ The omnipresence of the Web has already raised users’ expectations with respect to linking everything together: OPACS, A&I databases, e-journals and other full-text archives—even the whole Internet.

Links are sometimes treated as entities themselves, especially within the digital libraries. Separate link databases are flourishing, like SilverLinker, CrossRef and other commercially available services, together with a diversity of proprietary solutions. It should be noted that most of these are ‘closed’ or ‘static’ in some respect.

10 Web Services

The Web today (still) depends on us to use browsers to access information services, then to manually parse and analyze the displayed data, in order to identify the roads leading to the desired goals.

Now Web services are here. Web services are strictly organized and standardized Web applications that can be used by other network applications (not browsers) in order to perform a limited task. (These are commonly called business-to-business applications, as opposed to business-to-user.) In general, the use of Web services ought to offer us new ways of integrating and tailoring our information systems, better modularization and extended possibilities for the reuse of tools and services. A potentially interesting area for the application of Web services should be within linking, where separate link services can be accessed by other applications such as the library catalogs on the Web.

11 Reference Linking

Today, there is a huge demand for user-friendly methods for reference linking. This is the class of links that can be somewhat vaguely described as linking from metadata (reference, citation) to the full content.

The source may be a metadata record in a database or citation (more or less formally expressed) within some document. The target (full content) may be 'anything, anywhere' with a network identifier.

Some common examples of reference links:

- From an A&I database record to the full text;
- From a citation included in a document to the full text;
- From an OPAC record to an e-journal TOC with further linking possibilities;
- From an OPAC record to a network full-text manifestation of a document.

12 Static vs Dyamic Links

Most linking architectures are static, in the sense that the links are precomputed ('just in case,' 'a priori'), the target space is a controlled environment, and the links are more or less 'foolproof.' On the other hand, we might describe dynamic links in this way: These links are created 'a posteriori' (just in time), the target space need not be controlled, and the links are probabilistic (they might not work). Certainly dynamic link creation can include link verification, but this probably takes too long in most applications. (And it seems that automatic link verification can never be 100% reliable.)

As a real-world example of a static reference linking service, I can mention CrossRef:

- This linking service is operated by PILA (Publisher International Linking Association);
- CrossRef is implemented as a static link database;
- The link targets are DOIs (Digital Object Identifiers); and

- Access to the DOI resolution database (metadata → DOI) requires PILA membership.

13 Extended Service Links

Reference links usually target one specific copy of the full-content entity. But the user might rather need or prefer

- full content from another supplier;
- an OPAC holdings description;
- a copy ordering/ILL service;
- another metadata description/abstract;
- a book review or access to a net bookshop;
- a 'full Web' search.

These are often described as extended services.

It is obvious that every conceivable link is not appropriate to the user, because of

- diverse personal preferences (formats, delivery options, etc.);
- diverse institutional preferences;
- access restrictions, and/or
- temporary unavailability.

These and other parameters constitute the context of the user. The appropriateness of the link depends on this context.

14 Closed vs Open Linking

By closed links we understand links that are not context-sensitive:

- They might not work (because of access restrictions);
- They ignore the policy of the user's library; and
- They ignore the user's 'real' needs and preferences:

By contrast, we will use the term ‘open links’ for those that are context-sensitive. And furthermore, this means that open linking architectures support extended services

One well-known and pioneering implementation of such an open linking architecture is SFX (‘Special Effects’), which is now a part of the MetaLib product from Ex Libris.

15 OpenURL and Service Components

OpenURL is often considered to be a framework for implementing open linking. But strictly speaking, the OpenURL itself is just a standardized syntax for encoding metadata for a document into a URL “... to enable the transfer of the metadata from the information service to a service component that can provide context-sensitive services for the transferred metadata.” There may be several service components available, offered by different agencies or service suppliers. The OpenURL is presently under consideration as a NISO standard.

An OpenURL may look like this:

```
http://www.bibsys.no/ourl?sid=BIBSYS:ERL&issn=1234-
        5678&date=1998&volume=12&issue=2&spage=134
```

The different parts of this URL must comply with the syntactical requirements of the OpenURL standard.

16 Open Linking in Library Catalogs

The preceding discussion may appear to be targeted mostly towards digital libraries. But even in library catalogs, we can consider

- the use of open linking architectures;
- support of OpenURL;
- implementing separate link services;
- support of Web services.

We are seeing the first attempts to implement OpenURLs and separate link services in library catalogs. Web services will surely follow. It is my

hope that such measures may lead to more user-friendly and interoperable systems.

References

Functional Requirements for Bibliographic Records: Final Report. UBCIM Publications, New Series, vol. 19. Munchen: K.G. Saur, 1998. Available at http://www.ifla.org/VII/s13/frbr/frbr.pdf.

Gardner, Tracy. "An Introduction to Web Services." *Ariadne* 29 (2001). Available at http://www.ariadne.ac.uk/issue29/gardner/.

Hickey, Thomas B., Edward T. O'Neill, and Jenny Toves. "Experiments with the IFLA Functional Requirements for Bibliographic Records (FRBR)." *D-Lib Magazine* 8(9). Available at http://www.dlib.org/dlib/september02/hickey/09hickey.html.

Tillett, Barbara B. "Bibliographic Relationships". In *Relationships in the Organization of Knowledge*. Carol A. Bean and Rebecca Green, eds. Norwell, MA: Kluwer: 19–35.

Van de Sompel, Herbert, and Oren Beit-Arie. "Open Linking in the Scholarly Information Environment Using the OpenURL Framework." *D-Lib Magazine* 7 (3). Available at http://www.dlib.org/dlib/march01/vandesompel/03vandesompel.html.

Aalberg, Trond. "Navigating in Bibliographic Catalogs." In *Proceedings of the 6th European Conference on Research and Advanced Technology for Digital Libraries* (ECDL 2002). Springer Lecture Notes in Computer Science, vol. 2458: 238–250.